

# Synopse der Hohenheimer Metaanalysen zur Prognostizierbarkeit des Studienerfolgs und Implikationen für die Auswahl- und Beratungspraxis<sup>1</sup>

Benedikt Hell, Sabrina Trapmann & Heinz Schuler

*Um ein guter Student zu sein, muss man eine Leichtigkeit der Auffassung haben, Willigkeit, seine Kräfte auf all das zu konzentrieren, was einem vorgetragen wird; Ordnungsliebe, um das in den Vorlesungen Dargebotene schriftlich aufzuzeichnen und dann gewissenhaft auszuarbeiten.*

*All diese Eigenschaften fehlten mir gründlich!*

(Albert Einstein)

## 1 Einleitung

In jüngster Zeit ist ein stark ansteigendes Interesse an Auswahl- und Zulassungsverfahren für den Hochschulbereich zu verzeichnen. Ein besonderer Impuls ging von der Novelle des Hochschulrahmengesetzes im Jahre 1998 aus. Den Hochschulen wurden erheblich mehr Freiheiten bei der Zulassung eingeräumt und in zunehmendem Ausmaße nutzen sie die neu entstandenen Freiheiten dazu, Auswahlverfahren zu etablieren (Heine, Briedis, Didi, Haase & Trost, 2006).

Zur Klärung der bedeutsamen Frage, welche diagnostischen Verfahren geeignet sind, Studienerfolg zu prognostizieren, wurden in den vergangenen Jahrzehnten zahlreiche empirische Studien veröffentlicht. Eine geeignete Methode zur Integration einer großen Zahl von Einzelstudien, die ihre Erkenntnisse mittels vergleichbarer Forschungsdesigns gewinnen, stellt die so genannte Metaanalyse bzw. die Methode der Validitätsgeneralisierung dar (Hunter, Schmidt & Jackson, 1982). Unter Verwendung dieses Ansatzes bestimmte die Hohenheimer Arbeitsgruppe in den letzten Jahren die Prognosekraft von Schulnoten (Schuler, Funke & Baron-Boldt, 1990; Trapmann, Hell, Weigand & Schuler, 2007), fachspezifischen Studierfähigkeitstests (Hell, Trapmann & Schuler, 2007) und Auswahlgesprächen (Hell, Trapmann, Weigand & Schuler, 2007). Ziel dieses Beitrags ist die integrative Darstellung der drei aktuellen Hohenheimer Metaanalysen. Besonderer Wert wird dabei auf die Ableitung der für die Auswahl- und Beratungspraxis relevanten Schlussfolgerungen aus den Ergebnissen gelegt.

---

<sup>1</sup> Dieser Beitrag entstand innerhalb des Projekts „Eignungsdiagnostische Auswahl von Studierenden“, das im Rahmen des Aktionsprogramms „StudierendenAuswahl“ des Stifterverbands für die Deutsche Wissenschaft und der Landesstiftung Baden-Württemberg durchgeführt wurde.

## 2 Methode

Studienerfolg kann grundsätzlich sehr unterschiedlich operationalisiert werden. Im allgemeinen Sprachgebrauch versteht man unter „Erreichen des Studienziels“, dass ein einmal angefangenes Studium mit dem Hauptexamen abgeschlossen wird. Alternativ oder ergänzend bieten sich andere Indikatoren des Studienerfolgs an: Erreichte Noten in Zwischen- oder Abschlussprüfungen, erreichte Noten in einzelnen Lehrveranstaltungen, Studiendauer, Prüfungswiederholungen, Beurteilung des Studienerfolgs – beispielsweise durch Dozenten oder Kommilitonen – oder Angaben über die Zufriedenheit mit dem Studium. Da die weitaus meisten empirischen Studien zur Validierung von Auswahlverfahren das Erfolgskriterium Studiennoten heranziehen, wurde dieses Kriterium im Rahmen der Hohenheimer Metaanalysen zur Integration der Einzelbefunde verwendet.

Bei allen drei Metaanalysen wurde die auf Hunter et al. (1982) zurückgehende Methode der Validitätsgeneralisierung nach Raju, Burke, Normand und Langlois (1991) angewendet, die eine im Vergleich zu anderen Methoden genauere Schätzung der wahren Validität ( $\rho$ ) und ihrer Varianz liefert.

Im Ergebnisteil werden sowohl Konfidenz- als auch Kreditibilitätsintervalle berichtet. Das Kreditibilitätsintervall bezieht sich auf die Variabilität der Effekte über Studien hinweg. Hierzu wird die Varianz der korrigierten Validitäten herangezogen. Ist die untere Grenze des 90 %-Kreditibilitätsintervalls größer als Null, liegen 90 % der  $\rho$ -Verteilung über Null. Werden 100 % der Varianz durch die berücksichtigten Artefakte aufgeklärt, so entspricht der Kreditibilitätswert dem korrigierten Validitätskoeffizienten. Das Konfidenzintervall hingegen zeigt das Vertrauensintervall um  $\rho$  an. Hierzu wird der Stichprobenfehler herangezogen.

Es wurden Korrekturen von Artefaktquellen durchgeführt, die zu einer artifiziellen Verwässerung der beobachteten Effekstärkevarianz beitragen. Bei allen Korrelationen wurde eine Kontrolle des Stichprobenfehlers vorgenommen. Auch die Reliabilität des Kriteriums sowie Varianzeinschränkungen (Selektion der Studierenden durch das durchlaufene Auswahlverfahren, Selektion durch Studienabbruch) wurden berücksichtigt. Auf eine Korrektur der Reliabilität der Prädiktoren wurde verzichtet.

Die Studien enthielten keine verlässlichen Informationen zur Reliabilität der Studiennoten. Daher wurde für die Korrektur der Reliabilität des Kriteriums Durchschnittsnote der von Kuncel, Hezlett und Ones (2001, 2004) ermittelte Wert  $r_{tt} = .83$  verwendet. Für die Korrektur der Reliabilität von Einzelnoten wurde bei den referierten Metaanalysen der gemäß Spearman-Brown-Prophecy-Formula extrapolierte Wert von  $r_{tt} = .71$  angenommen.

## 3 Ergebnisse

Die drei diesem Beitrag zugrunde liegenden Metaanalysen integrieren Studien, die international (Prädiktor Interview), in Europa (Prädiktor Schulnoten) oder im deutschsprachigen Raum (Prädiktor Studierfähigkeitstests) seit 1980 publiziert worden

sind<sup>2</sup>. Tabelle 1 hält einen Überblick über die Anzahl der unabhängigen Stichproben und die Größen der jeweiligen Gesamtstichprobengröße differenziert nach untersuchtem Prädiktor bereit.

**Tabelle 1:** Anzahl der Validitätskoeffizienten und Stichprobengröße nach Prädiktor

	<b>N</b>	<b>k</b>
Durchschnittsnote	48.178	53
Fachspezifischer Studierfähigkeitstest	45.091	36
Interview, alle Formen	4.305	44
Interview, unstrukturiert	2.136	14
Interview, strukturiert	2.169	30

*Anmerkungen:* N = Größe der Gesamtstichprobe; k = Anzahl der unabhängigen Stichproben

Die Tatsache, dass die Gesamtstichprobengrößen der Prädiktoren Schuldurchschnittsnote und fachspezifischer Studierfähigkeitstest erheblich über der Gesamtstichprobengröße des Prädiktors Interview liegen, ist darauf zurückzuführen, dass es sich bei den beiden erstgenannten Verfahren um Auswahlkomponenten handelt, die auch bei großen Bewerberzahlen angewendet werden, wohingegen Auswahlgespräche aufgrund des zu leistenden Aufwands nur bei überschaubaren Bewerbergruppen zum Einsatz kommen.

### 3.1 Schulnoten

Sowohl die aktuelle Metaanalyse (Trapmann et al., 2007), deren Ergebnisse in komprimiertester Form in Tabelle 2 dokumentiert sind, als auch die Metaanalyse von Schuler et al. aus dem Jahr 1990 belegen die ausgezeichnete Validität der Durchschnittsnote der Hochschulzugangsberechtigung. Schulnoten stellen den besten Einzelprädiktor von Studienerfolg im Sinne von Studiennoten dar. Die Analysen zeigen weiterhin, dass ein Anstieg in der Validität der Schulnoten von der 1990er Metaanalyse zur aktuellen Metaanalyse und auch innerhalb des Zeitraums der aktuellen Metaanalyse zu verzeichnen ist.

Das Herkunftsland der Studie moderiert ebenso wie das Studienfach und der Studienabschnitt, in dem der Studienerfolg gemessen wurde, die Validität der Schulnoten: Die

<sup>2</sup> Eine Ausnahme stellen Studien zur Prognosekraft der Schulnoten aus dem deutschsprachigen Raum dar, die ab dem Publikationsjahr 1985 aufgenommen wurden, um eine Überschneidung mit der Analyse von Schuler et al. (1990) zu vermeiden.

**Tabelle 2:** Die Validität der Durchschnittsnote der Hochschulzugangsberechtigung für Studiennoten

	$\rho$	%	95 % KI <sub>FE</sub>	95 % KI <sub>RE</sub>	90 % KW
Durchschnittsnote	.517	12.4	[.507, .527]	[.489, .545]	.393

*Anmerkungen:*  $\rho$  = Mittlere korrigierte Validität; % = Prozentualer Anteil der Gesamtvarianz, der durch die Artefakte erklärt wird; 95 % KI<sub>FE</sub> = Konfidenzintervall mit  $p=95\%$  um  $\rho$ , Modell: Feste Effekte; 95 % KI<sub>RE</sub> = Konfidenzintervall mit  $p=95\%$  um  $\rho$ , Modell: Zufällige Effekte; 90 % KW = 90 %-Kreditabilitätswert für die Verteilung der wahren Werte (untere Grenze)

Analysen zeigen, dass die deutschen Abiturnoten besonders aussagekräftig sind, sich die Studienfächer Mathematik und Naturwissenschaften am besten prognostizieren lassen und weiterhin, dass frühe Studienleistungen (Grundstudium, Bachelor, vorklinischer Abschnitt) im Vergleich zu späten Studienleistungen (Hauptstudium, Master, klinischer Abschnitt) enger mit den Schulleistungen zusammenhängen.

### 3.2 Fachspezifische Studierfähigkeitstests

Durch eine etwas geringere aber immer noch sehr günstige Validität zeichnen sich fachspezifische Studierfähigkeitstests aus (vgl. Tabelle 3). Diese Verfahrensklasse ist in der Lage, 22,8 % der Kriterienvarianz aufzuklären. Als bedeutsamer Moderator der Validität stellte sich insbesondere das Studienfach heraus. Die höchste Validität wird im Studienfach Humanmedizin ( $\rho = .507$ ), die geringste Validität im Studienfach Wirtschaftswissenschaften ( $\rho = .325$ ) erreicht.

**Tabelle 3:** Die Validität von fachspezifischen Studierfähigkeitstests für Studiennoten

	$\rho$	%	95 % KI <sub>FE</sub>	95 % KI <sub>RE</sub>	90 % KW
Fachspezifische Studierfähigkeitstests	.478	18.7	[.468, .489]	[.454, .503]	.392

*Anmerkungen:* vgl. Tabelle 2

Fachspezifische Studierfähigkeitstests (FSFT) stehen in Konkurrenz zu Schulnoten, die traditionell zur Studierendenauswahl eingesetzt werden. Daher liegt ein direkter Vergleich der Validitäten nahe. Ein Vergleich der Validitäten von Schulnoten und Studierfähigkeitstests zeigt, dass Schulnoten für alle Studienfächer der stärkere Prädiktor sind: Für das Studienfach Humanmedizin weisen die Schuldurchschnittsnoten eine Validität von .551 (vs. FSFT: .507), für Veterinärmedizin von .576 (vs. FSFT: .431), für Wirtschaftswissenschaften von .484 (vs. FSFT: .325) und für Zahnmedizin von .375

(vs. FSFT: .353) auf. Ob ein gemeinsamer Einsatz von Studierfähigkeitstests und Schulnoten indes sinnvoll ist, hängt vom Ausmaß der inkrementellen Validität ab. Diese wiederum ergibt sich aus der eigenständigen Validität der Prädiktoren und aus der Korrelation der beiden Prädiktoren untereinander. Sind die beiden Prädiktoren unkorreliert, ist eine hohe inkrementelle Validität zu erwarten. Je stärker die beiden Prädiktoren korreliert sind, sie also Ähnliches messen, desto geringer fällt die inkrementelle Validität aus.

Nimmt man die berichteten Validitätskoeffizienten für Schulnoten und Studierfähigkeitstests als Grundlage, so sind die in Tabelle 4 in der rechten Spalte aufgeführten Validitätskennwerte in Abhängigkeit von den in der linken Spalte aufgeführten Interkorrelationen der beiden Prädiktoren bei einem gemeinsamen Einsatz von Schulnoten und Studierfähigkeitstest zu erwarten. Dabei stellen die angegebenen Validitätskennwerte im Falle eines kombinierten Einsatzes von Schulnoten und Studierfähigkeitstests eine optimistische Schätzung dar, da die beiden Prädiktoren mittels multipler Regressionen optimal gewichtet wurden. Solche Steigerungen der Validität der Schulnoten sind nur unter idealen Bedingungen (regressionsanalytische Gewichtung, sorgfältige Kreuzvalidierung) und bei einer stärkeren Gewichtung der Schulnoten zu erwarten.

**Tabelle 4:** Zu erwartende inkrementelle Validität durch den gemeinsamen Einsatz von Studierfähigkeitstests und Schulnoten

<b>Interkorrelation zwischen Test und Schulnote</b>	<b>Gemeinsame Validität bei optimaler Gewichtung</b>
.00	.71
.20	.65
.40	.60
.60	.56

*Anmerkungen:* Angenommene eigenständige Validitätskoeffizienten: Schulnoten = .52; Studierfähigkeitstest = .48

Über das bislang im deutschsprachigen Raum am gründlichsten erforschte Instrument, den Test für medizinische Studiengänge (TMS), berichten Trost et al. im 1998 erschienenen Abschlussbericht, dass die Korrelation zwischen TMS und der „Durchschnittsnote der Schule im betrachteten Zeitraum minimal .37 und maximal .48 beträgt“ (S. 56). Legt man diese Unter- und Obergrenze der Interkorrelation und die berichteten eigenständigen Validitäten der beiden Prädiktoren zugrunde, so ergibt sich ein geschätzter Wertekorridor für die Validität eines kombinierten Verfahrenseinsatzes von Schulnoten und Studierfähigkeitstests, der von .58 (Interkorrelation Prädiktoren: .48) bis .61 (Interkorrelation Prädiktoren: .37) reicht. Für die nachfolgenden Überlegungen wird von einem mittleren Validitätskoeffizienten im Falle eines kombinierten Verfahrenseinsatzes von .60 ausgegangen.

Im Rahmen der psychologischen Diagnostik entwickelte Nutzenmodelle erlauben die genaue Bestimmung von Trefferquoten (Anteil der geeigneten Studierenden unter den Ausgewählten<sup>3</sup>) in Abhängigkeit von der Validität der Auswahlverfahren und sind somit in der Lage, genaue Informationen über die Sinnhaftigkeit von Verfahrenskombinationen zu liefern. Die Trefferquote eines Auswahlverfahrens ist neben der Validität von der Basisrate (Anteil der Geeigneten unter den Bewerbern) und von der Selektionsquote (Prozentsatz der angenommenen Bewerber) abhängig (vgl. Taylor & Russel, 1939). Eine genauere Vorstellung über den Nutzen von Studierfähigkeitstests gewinnt man durch einen Vergleich der zu erwartenden Trefferquoten für die eigenständige Validität von Schulnoten (.52) und der gemeinsamen Validität von Schulnoten und Studierfähigkeitstests (.60) in Abhängigkeit von verschiedenen Selektionsquoten und Basisraten (vgl. Tabelle 5).

**Tabelle 5:** Trefferquoten bei kombiniertem Einsatz von Studierfähigkeitstest und Abiturdurchschnittsnote (Validität: .60) gegenüber alleiniger Berücksichtigung der Abiturdurchschnittsnote (Validität: .52)

Validität	Selektionsquote									
	.05	.15	.25	.35	.45	.55	.65	.75	.85	.95
<b>Basisrate: .20</b>										
<b>.52</b>	.60	.48	.42	.37	.33	.30	.28	.25	.23	.21
<b>.60</b>	.68	.54	.46	.40	.36	.32	.29	.26	.23	.21
<b>Δ</b>	.08	.06	.04	.03	.03	.02	.01	.01	.00	.00
<b>Basisrate: .40</b>										
<b>.52</b>	.83	.73	.68	.63	.59	.55	.52	.48	.45	.42
<b>.60</b>	.88	.79	.72	.67	.62	.58	.53	.49	.46	.42
<b>Δ</b>	.05	.06	.04	.04	.03	.03	.01	.01	.01	.00
<b>Basisrate: .60</b>										
<b>.52</b>	.94	.89	.85	.81	.78	.75	.72	.69	.66	.62
<b>.60</b>	.96	.92	.88	.85	.81	.78	.74	.71	.67	.63
<b>Δ</b>	.02	.03	.03	.04	.03	.03	.02	.02	.01	.01

Anmerkung: Δ = Differenz Trefferquote

<sup>3</sup> Als geeignete *Studierende* werden in diesem Zusammenhang diejenigen Studierenden aufgefasst, die ihr Studium mit Erfolg abschließen. Diese kriterienbezogene Definition unterscheidet sich von der gesetzgeberischen Definition der Studieneignung, nach der jeder Bewerber mit einer gültigen Hochschulzugangsberechtigung für ein Studium geeignet ist. Eine Auswahl ist nach derzeitiger Gesetzeslage nur dann vorgesehen, wenn die Anzahl der Bewerber die Anzahl der Studienplätze übersteigt.

Mittels der tabellierten Trefferquoten können Universitäten individuell entscheiden, ob die Berücksichtigung von Testergebnissen zusätzlich zu Schulnoten für sie in Frage kommt. Generell gilt, dass ein Testeinsatz eher bei strengen Selektionsquoten sinnvoll ist. Auch aus praktischen Gründen wird ein zusätzlich zu Schulnoten eingesetzter Test bei strengen Selektionsquoten interessant, denn gerade bei außerordentlich beliebten Studiengängen erreicht der Numerus clausus unter Umständen ein so hohes Niveau, dass viele Bewerber auch durch Wartezeiten oder sonstige Boni die Zulassungsbedingungen nicht erfüllen können. Der Einsatz eines weiteren Verfahrens eröffnet dieser Bewerbergruppe eine zusätzliche Chance.

Eine endgültige Nutzenabwägung im Sinne einer Kostenkalkulation ist im vorliegenden Kontext nicht möglich, da auf mehreren Seiten sowohl Vorteile als auch Nachteile mit einer Testverwendung verbunden sind. Aus Sicht der *Universitäten* verursachen ungeeignete Studierende dadurch Kosten, dass sie ihr Studium abbrechen oder überdurchschnittlich lange studieren und somit Studienplätze blockieren. Diese Kosten lassen sich beziffern und es ist möglich, mit den aufgeführten Trefferquotendifferenzen die Einsparungen zu berechnen, die die Berücksichtigung von Tests bei der Auswahl einbringen würde. Diesem Betrag stehen die Kosten für die Testentwicklung, Durchführung und Auswertung gegenüber, die bei einem bundesweit einheitlichen Test gegebenenfalls von den Bewerbern getragen werden könnten. Solche Nutzenüberlegungen greifen allerdings zu kurz, da sie lediglich die Perspektive der Hochschulen einnehmen. Ihr sind andere Gesichtspunkte an die Seite zu stellen: Sind die Verfahren für die *Bewerber* „nützlich“? Sind z. B. die Trefferquoten bei allen Subgruppen ähnlich (Fairness)? Unter bestimmten Umständen sind Verzerrungen (Fehler 2. Art) zu erwarten und hierdurch wären aus Sicht der Bewerber erhebliche „Kosten“ im Sinne von verstellten Entwicklungsmöglichkeiten zu erwarten (vgl. Raatz, 1978). Aus Sicht der *Gesellschaft* wiederum stellt sich die Regelung der Zulassung zu einem Hochschulstudium eher als Zuordnungsentscheidung denn als Selektionsentscheidung dar. Ziel sollte es also sein, für jede Person den passendsten Studienplatz zu finden und nicht eine Bestenauswahl zu betreiben.

### 3.3 Interviews

Die prognostische Validität von Auswahlgesprächen für Studiennoten ist entgegen der häufig von Laien geäußerten Meinung vergleichsweise gering. Es kann eine niedrige, aber immerhin durch das Kreditabilitätsintervall und auch das Konfidenzintervall abgesicherte positive Validität generalisiert werden (vgl. Tabelle 6).

Den bedeutendsten Moderator der Validität von Auswahlgesprächen stellt der Strukturierungsgrad der Interviews dar; strukturierte Interviews erreichen immerhin eine Validität von  $\rho = .205$ , woraus eine Varianzaufklärung hinsichtlich des Erfolgskriteriums Studiennoten von 4,2 % resultiert. Die Validität unstrukturierter Interviews hingegen ist sehr spärlich; sie sind als Prädiktoren für Studiennoten unbrauchbar.

Von großer Bedeutung ist die inkrementelle Validität von Auswahlgesprächen gegenüber anderen Zulassungsverfahren. Die inkrementelle Validität zu Schulnoten und

zu Studierfähigkeitstests interessiert besonders, da diese beiden Verfahren die höchste eigenständige prognostische Validität erbringen und zudem am häufigsten eingesetzt werden.

**Tabelle 6:** Ergebnisse der Metaanalyse differenziert nach Form des Interviews

	$\rho$	%	95 % KI <sub>FE</sub>	95 % KI <sub>RE</sub>	90 % KW
Interview, alle Formen	.157	90.1	[.118, .196]	[.116, .198]	.101
Interview, unstrukturiert	.109	100.0	[.046, .171]	[.054, .164]	.109
Interview, strukturiert	.205	89.4	[.154, .256]	[.151, .259]	.142

Anmerkungen: vgl. Tabelle 2

**Tabelle 7:** Inkrementelle Validität von Auswahlgesprächen für das Erfolgskriterium Studiennoten

Prädiktor/Modell	Prädiktorkombination	$\beta$	R	R <sup>2</sup>
Inkrementelle Validität gegenüber Schulnoten				
Annahme: $r_{\text{Schulnoten, Interview}} = .16$				
Modell 1	Schulnoten	.517	.517	.267
Modell 2	Schulnoten	.505	.522	.273
	Interview	.076		
Inkrementelle Validität gegenüber fachspezifischen Studierfähigkeitstests				
Annahme: $r_{\text{FSFT, Interview}} = .12$				
Modell 1	Studierfähigkeitstest	.478	.478	.228
Modell 2	Studierfähigkeitstest	.466	.489	.239
	Interview	.103		

Für die Kalkulation der inkrementellen Validität wurden von Hell et al. (2007) die Interkorrelation zwischen Auswahlgesprächen und den beiden anderen Verfahren metaanalytisch ermittelt. Anschließend wurden mittels multipler Regressionen die inkrementellen Validitäten berechnet. Wie in Tabelle 7 zu sehen ist, kann durch den zusätzlichen Einsatz von Auswahlgesprächen die eigenständige Validität der Schulnoten von .517 auf .522 und die eigenständige Validität von Studierfähigkeitstests von .478 auf .489 nur marginal gesteigert werden.



## 4 Implikationen für die Auswahl- und Beratungspraxis

Die Hohenheimer Metaanalysen belegen, dass das Studienerfolgskriterium Studiennoten vergleichsweise gut prognostiziert werden kann. Eine Kombination der beiden Verfahren Schulnoten und Studierfähigkeitstests lässt eine Varianzaufklärung am Kriterium Studiennoten von 36 % erwarten. Gleichwohl sind Zulassungsverfahren für jede Auswahl-situation individuell anzupassen. Dies gilt insbesondere im Hinblick auf die im Zuge des Bologna-Prozesses geschaffene zweistufige Studienstruktur.

Die prädiktive Validität von *durchschnittlichen Schulabschlussnoten* liegt an der Spitze der Einzelprädiktoren zur Vorhersage des Studienerfolgs. Ähnlich hohe Werte erreichen nur noch Studierfähigkeitstests. Die Detailanalysen von Trapmann et al. (2007) zeigen weiterhin, dass Durchschnittsnoten wesentlich valider sind als Einzelnoten. Dieses Ergebnis überrascht nicht, sondern bestätigt die bereits in der Metaanalyse von Schuler et al. (1990) berichteten Ergebnisse. Die Autoren führten damals an, dass die Durchschnittsnote aufgrund der Aggregation die messtheoretischen Anforderungen an einen Prädiktor besser als die einzelnen Fachnoten erfüllt. Tatsächlich handelt es sich bei den deutschen Abiturnoten um Urteile, die über zwei Jahre hinweg von mehreren Beurteilern in Prüfungen verschiedener Modalitäten (schriftliche Klausuren und Tests, mündliche Mitarbeit, mündliche Abiturprüfung) gebildet werden. Darüber hinaus spiegeln die Schulabschlussnoten Fähigkeiten und Fertigkeiten wider, die sowohl für gute Schulnoten als auch für gute Studiennoten relevant sind, wie z. B. kognitive Leistungsfähigkeit, Lernbereitschaft, Leistungsmotivation, Fleiß und sprachliche Ausdrucksfähigkeit.

Da Schuldurchschnittsnoten die höchste Einzelvalidität aufweisen und sie zudem mit vergleichsweise geringem Aufwand durch die Hochschulen ermittelbar sind, sollten sie aus eignungsdiagnostischer Sicht Bestandteil von Auswahlverfahren sein. Das Hochschulrahmengesetz formuliert in Übereinstimmung mit diesen Erkenntnissen, dass der Abiturdurchschnittsnote ein „maßgeblicher Einfluss“ eingeräumt werden muss.

Wengleich Schulnoten auch für Noten in Masterstudiengängen valide sind (Trapmann et al., 2007), würde eine Berücksichtigung dieser zeitlich weit zurückliegenden und inhaltlich nicht in direktem Bezug zum Masterstudium stehenden Leistungen bei der Zulassung auf Akzeptanzprobleme bei den Bewerbern stoßen. Gegen die ausschließliche Verwendung der Noten des Bachelorabschlusses zur Auswahl der Master-Aspiranten wiederum spricht die teilweise erhebliche Heterogenität der für die Masterabschlüsse jeweils einschlägigen Bachelorabschlüsse, die sich mindernd auf die Validität der Bachelornoten auswirkt. Es empfiehlt sich daher, Bachelornoten durch Verfahren zur ergänzen, die gleiche Anforderungen an die Bewerber stellen. In Anbetracht der günstigen Validität ist dabei an Studierfähigkeitstests und unter bestimmten Umständen (vgl. Überlegungen von Trapmann, Hell & Schuler, in diesem Band) an Auswahlgespräche zu denken.

Eine interessante, bislang wenig beachtete Forschungsfrage stellt die internationale Vergleichbarkeit der Schulabschlüsse und Schulnoten dar. Da für sämtliche Studienplatzbewerber aus dem europäischen Raum eine Gleichbehandlung und eine Berück-

sichtigung der Noten der Hochschulzugangsberechtigung bei Auswahlverfahren vorgeschrieben sind, wäre eine Vergleichbarkeit der Schulabschlüsse und Schulnoten wünschenswert. Angesichts unterschiedlicher Bildungssysteme und unterschiedlicher Abiturientenquoten ist eine Gleichartigkeit der Notenniveaus aber nicht zu erwarten. Eine direkte lineare Transformation der Noten der ausländischen Hochschulzugangsberechtigungen, wie sie derzeit in Deutschland in Form der sogenannten modifizierten bayrischen Formel vorgeschrieben ist (Beschluss der Kultusministerkonferenz vom 15. 03. 1991 i. d. F. vom 18. 11. 2004), leistet hier keine Abhilfe, da zwar Mittelwerteffekte ausgeglichen werden, aber die unterschiedlichen Verteilungsformen keine Berücksichtigung finden. Zielführender erscheint eine Prozentrang-Normierung der Schulnoten (Hell, 2007).

*Fachspezifische Studierfähigkeitstests* verfügen über eine substanzielle eigenständige Validität und sind zudem in der Lage, die Prognosekraft von Schulnoten durch eine gemeinsame Berücksichtigung dieser beiden Zulassungskriterien zu steigern. Letzteres gilt nicht für alle Auswahl-situationen. Eine inkrementelle Validität von Studierfähigkeitstests ist unter definierten Bedingungen zu erwarten. Entscheidend sind die Ausprägungen der drei Parameter Basisrate, Selektionsquote und Erfolgsquote, die jede Universität hochschulspezifisch ermitteln und zur fundierten Entscheidung, ob ein Studierfähigkeitstest zur Zulassung eingesetzt werden soll, heranziehen kann. Insbesondere bei der Zulassung zu Masterprogrammen sollte aus den oben angeführten Gründen über den Einsatz von Studierfähigkeitstests nachgedacht werden.

Aus theoretischen wie pragmatischen Gründen bietet ein bundesweit einheitlich angebotener Test gegenüber hochschulspezifischen Verfahren einige Vorteile. So wäre es den Bewerbern möglich, den Test einmal zu absolvieren und sich anschließend bei verschiedenen Hochschulen mit einem Zertifikat zu bewerben; der durch mehrfache Testteilnahmen an mehreren Hochschulen entstehende, für Bewerber und Hochschulen erhebliche Aufwand würde insgesamt reduziert. Gleichzeitig könnte die Testentwicklung und -evaluation zentral erfolgen. Dadurch böte sich die Gelegenheit, kostengünstige und qualitativ hochwertige Verfahren bereitzustellen und ihre Qualitätssicherung auf Dauer zu gewährleisten.

Die Ergebnisse der Hohenheimer Metaanalysen zeigen, dass *Auswahlgespräche* zur Prognose von Studiennoten kaum geeignet sind. Die zu erwartenden inkrementellen Zuwächse der Prognosekraft durch die gemeinsame Verwendung von Auswahlgesprächen und Schulnoten bzw. Studierfähigkeitstests sind vom Betrag her so gering, dass der zusätzliche Einsatz von Interviews zur Prognose von Studiennoten nicht als sinnvoll erachtet werden kann, wenn bereits valide Informationen wie Schulnoten oder Ergebnisse von Studierfähigkeitstests vorliegen. Die empirische Evidenz steht somit im Widerspruch zu der subjektiven Einschätzung der Interviewer (Wagoner & Gray, 1979) und auch der Bewerber (Hell & Schuler, 2005), dass Interviews besonders aussagekräftig sind.

Hinsichtlich anderer proximaler Erfolgskriterien wie *Studiendauer*, *Zufriedenheit* und *Studienabbruch* oder *Berufserfolg* als distales Studienerfolgskriterium können aus den Hohenheimer Metaanalysen keine Ableitungen gemacht werden. Vorausgesetzt, die in

der beruflichen Eignungsdiagnostik gewonnenen Erkenntnisse hinsichtlich der positiven Wirkung realistischer Tätigkeitsinformation auf Fluktuationsraten (Phillips, 1998) können auf den Studienkontext übertragen werden, sollten zumindest Interviews, in denen Bewerber über Studieninhalte und -bedingungen informiert werden, für das Studien-erfolgskriterium (unterlassener) Studienabbruch prognostisch valide sein.

Die günstige Prognosekraft strukturierter Interviews für Berufserfolg hingegen ist unstrittig (zusammenfassend Schuler, 2002). Eine auf den Berufserfolg ausgerichtete Auswahl von Studierenden ist daher in Fächern denkbar, mit denen ein klar definiertes Berufsbild verbunden ist (z. B. Lehramtsstudiengänge; vgl. Trapmann et al., in diesem Band).

Auch bei der Zulassung zu Masterstudiengängen kommt die Verwendung von Auswahlgesprächen aus mehreren Gründen in Betracht. Zum einen sind mit den meisten Master-Programmen im Vergleich zu Bachelor-Studiengängen homogenere Berufs- und Karriereprofile verbunden, sodass eine Auswahl der Bewerber hinsichtlich des zu erwartenden beruflichen Erfolgs möglich ist. Zum anderen verfügen Interessenten für Master-Studiengänge über eine im Vergleich zu Bachelor-Bewerbern längere und somit aussagekräftigere Vita. Auswahlgespräche, die sowohl die definierten beruflichen Anforderungen als auch die vorhandenen biografischen Informationen sorgfältig abbilden, führen nicht nur einer hohen Akzeptanz bei Bewerbern und Verwendern, sondern können andere Auswahlelemente bei der Auswahl von Master-Studierenden sinnvoll ergänzen.

Auswahlgespräche im Rahmen der Hochschulzulassung müssen nicht zwangsläufig als reines Selektionsinstrument angesehen werden. Vielmehr stellt sich die Zulassungsregelung zum Hochschulstudium aus gesellschaftlicher Perspektive wie bereits angesprochen eher als Zuordnungsentscheidung denn als Selektionsentscheidung dar. In Interviews können für abgelehnte Bewerber Platzierungsempfehlungen entwickelt werden und bei Bedarf kann auf Beratungsangebote hingewiesen werden. Inwiefern Auswahlgespräche solche Zusatzfunktionen erfüllen sollen, ist bei Überlegungen, diese Verfahrensklasse einzuführen, zu berücksichtigen.

Ausgangspunkt für die Entwicklung von Zulassungsverfahren stellt die *Definition der Ziele* dar: Steht die Information und Beratung der Bewerber im Vordergrund? Soll das Niveau der Studienanfänger angehoben werden? Wird die Verringerung der Studienabbruchquote angestrebt? Aufbauend auf dieser Zieldefinition erfolgt die Analyse der Studien- und ggf. Berufsanforderungen. Zu diesem Zweck können *Anforderungsanalyseverfahren*, die im Rahmen der beruflichen Eignungsdiagnostik entwickelt wurden, in den vorliegenden Kontext transferiert werden (vgl. Hell, Ptok & Schuler, 2007). Im Ergebnis liefern diese Verfahren Listen von Anforderungen, die gemeinsam mit der Zieldefinition Grundlage der Verfahrenserstellung sein können.

Auf diese Weise entsteht ein Auswahl- und Beratungssystem, das auf die spezifischen Bedürfnisse der jeweiligen Universität zugeschnitten ist. Angesichts der Komplexität des Konstrukts Studieneignung und der häufig heterogenen Ziele, die mit Auswahl- und Beratungssystemen verfolgt werden, und nicht zuletzt aufgrund seiner psychometrischen Überlegenheit ist ein *multimethodales Vorgehen* empfehlenswert.

## Literatur

- Heine, C., Briedis, K., Didi, H.-J., Haase, C. & Trost, G. (2006). *Bestandsaufnahme von Auswahl- und Eignungsfeststellungsverfahren beim Hochschulzugang in Deutschland und ausgewählten Ländern*. Hannover: HIS-Kurzinformation A 3/2006.
- Hell, B. (2007). Die Prozentrangnormierung als Alternative zum bisherigen Umrechnungsverfahren ausländischer Sekundarabschlussnoten für die Verwendung in den Zulassungsverfahren deutscher Hochschulen. *Hochschulmanagement*, 4, 93–99.
- Hell, B., Ptok, C. & Schuler, H. (2007). Methodik zur Ermittlung und Validierung von Anforderungen an Studierende (MEVAS): Anforderungsanalyse für das Fach Wirtschaftswissenschaften. *Zeitschrift für Arbeits- und Organisationspsychologie*, 51, 88–95.
- Hell, B. & Schuler, H. (2005). Verfahren der Studierendenauswahl aus Sicht der Bewerber. *Empirische Pädagogik*, 19, 361–376.
- Hell, B., Trapmann, S. & Schuler, H. (2007). Eine Metaanalyse der Prognosekraft von Studierfähigkeitstests. *Empirische Pädagogik*, 21, 251–270.
- Hell, B., Trapmann, S., Weigand, S. & Schuler, H. (2007). Die Validität von Auswahlgesprächen im Rahmen der Hochschulzulassung – eine Metaanalyse. *Psychologische Rundschau*, 58, 93–102.
- Hunter, J., Schmidt, F. & Jackson, G. (1982). *Meta-analysis: Cumulating research findings across studies*. London: Sage.
- Kuncel, N. R., Hezlett, S. A. & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: Implications for graduate student selection and performance. *Psychological Bulletin*, 127, 162–181.
- Kuncel, N. R., Hezlett, S. A. & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, 86, 148–161.
- Phillips, J. M. (1998). Effects of realistic job previews on multiple organizational outcomes: A meta-analysis. *Academy of Management Journal*, 41, 673–690.
- Raatz, U. (1978). Was können Tests bei Entscheidungsproblemen leisten? *Lernzielorientierter Unterricht*, 2, 30–37.
- Raju, N. S., Burke, M. J., Normand, J. & Langlois, G. M. (1991). A new meta-analytic approach. *Journal of Applied Psychology*, 76, 432–446.
- Schuler, H. (2002). *Das Einstellungsinterview*. Göttingen: Hogrefe.
- Schuler, H., Funke, U., & Baron-Boldt, J. (1990). Predictive validity of school grades: A meta-analysis. *Applied Psychology: An International Review*, 39, 89–103.
- Taylor, H. C. & Russel, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. *Journal of Applied Psychology*, 23, 565–578.
- Trapmann, S., Hell, B., Weigand, S. & Schuler, H. (2007). Die Validität von Schulnoten zur Vorhersage des Studienerfolgs – eine Metaanalyse. *Zeitschrift für Pädagogische Psychologie*, 21, 11–27.
- Trost, G., Blum, F., Fay, E., Klieme, E., Maichle, U., Meyer, M. & Nauels, H.-U. (1998). *Evaluation des Tests für medizinische Studiengänge (TMS): Synopse der Ergebnisse*. Bonn: Institut für Test- und Begabungsforschung.
- Wagoner, N. E. & Gray, G. T. (1979). Report on a survey of program directors regarding selection factors in graduate medical education. *Journal of Medical Education*, 54, 445–452.